Contents

1 Cognitivism and Internalism

> Modern philosophy has never been able to quite shake off the Cartesian idea of the mind, as something that "resides," – such is the term, – in the pineal gland. Everybody laughs at this nowadays, and yet everybody continues to think of mind in this same general way, as something within this person or that, belonging to him and correlative to the real world.
>
> – Charles Sanders Peirce

1.1 Introduction

Cognitivism represents the major shift in the study of cognition after behaviourism and underpins the main theories and methodologies of cognitive science. In contrast to behaviourism, which focuses on observable behaviour, cognitivism posits internal representations. The explanatory focus turns to the processing of these representations to explain cognitive phenomena such as memory and is also used to explain observable behaviour. Cognition is simply defined as the processing of representations.

This claim is usually coupled to an assumption about where the representations are located and I will call this assumption "cognitive internalism." The assumption amounts to this: cognitive processes, whether computational or otherwise, occur inside the head. Hence, we find that cognitive processes and representations are "internal" and that they do not depend upon anything "external" to the cogniser. In other words, cognitive vehicles and processes are individuated independently of what goes on in the "external environment" of the cogniser – except perhaps for their content. A methodological moral follows: the study of cognitive processes should appeal only to what is inside the head. The point of this chapter is to show that cognitivism does not have to be an internalist doctrine. It is quite consistent with cognitive integration, which allows that some manipulations of external representational vehicles are cognitive processes. The internalist version of cognitivism is based upon some dubious metaphysical assumptions which nobody has quite been able to shake philosophers out of.[1] The direction of argument is simple; cognitivism is consistent with integration despite certain cherished assumptions about causal capacities and computational systems.

I will consider two arguments in favour of internalism that display the background metaphysical assumptions in question: Fodor's (1987) argument from causal capacities and, what I shall call, the computational argument (Fodor 1980, Segal 1989, Egan 1992). The causal capacities argument runs like this: (1) Science taxonomises by causal capacities (what an entity can cause). (2) The causal capacities an entity has supervened upon its intrinsic causal properties. (3) In the psychological case, the causal properties supervene upon local neural structure. (4) Therefore, cognitive capacities supervene upon intrinsic properties of the individual. Looked at in this way, the argument is a defence of mind–brain supervenience and I shall take it to be an argument in favour of cognitive vehicles and processes being located in the head. The argument is unsound if premises 2 and 3 are false. There are plenty of examples across the sciences that show premise 2 to be false. Cognitive integrationists think that premise 3 is false. The argument is important because it depends upon several metaphysical assumptions which drive an internalist reading of cognitivism. The primary metaphysical assumption is that only intrinsic physical properties are truly causal, they are the only properties that figure in genuinely causal generalisations. This assumption just turns out to be false in many of the sciences, including biology and the social sciences. Contemporary counter-arguments to cognitive integration (Adams and Aizawa 2001, 2007, Rupert 2004, 2007) depend upon this assumption in their appeal against genuine causal generalisations in psychology that involve anything other than intrinsic properties of the individual.

Wilson (1995, p. 64) gives us a version of the computational argument: (A) The sciences of cognition taxonomically individuate mental processes only qua computational processes. (B) The computational states and processes that an individual instantiates supervene on the intrinsic states of that individual. Therefore, (C) The sciences of cognition individuate states and processes that supervene on the intrinsic physical states of the individual who instantiates those states and processes. I will take the computational argument to be an argument in favour of the individuative independence of cognitive vehicles from the external environment. This argument is unsound if premise B is false. Clearly, premise A is false if not all cognitive processes are computational. Premise B is false if some cognitive processes are not intrinsic. I will not directly be arguing against A, I will be arguing for the conclusion that cognitive processes and vehicles are not intrinsic in the individualist sense.

The computational argument is important because the idea that a natural (or wide) psychology would be too difficult is a direct ancestor of Adams and Aizawa's complaint that such a science would amount to a disconnected motley (this complaint will be dealt with in Chapter 3).

Before turning to these arguments, I outline some of the terminology that will be retained from cognitivism, because cognitive integration is quite compatible with cognitivism once the internalist assumption has been removed from it.

1.2 What is a cognitive process and what is a cognitive system?

Most philosophers and cognitive scientists take cognition to be a clump of mental acts or processes that come under broad headings such as remembering, perceiving, learning and reasoning. Identifying what makes a process cognitive is more difficult. In a recent paper critical of the extended mind, processes that exhibit the mark of the cognitive are identified as those that involve representations with non-derived [intrinsic] content (Adams and Aizawa 2001). However, it is not only notoriously difficult to specify just what intrinsic content is supposed to be (Hutto 1999, Dennett 1990, Mendola 2003), but also the definition looks to be unduly restrictive (I shall have more to say about this condition in Chapter 4). Furthermore, we do not find cognitive scientists providing definitions of the "mark of the cognitive" as a preliminary to their empirical investigations.[2]

In general, there is no real agreement in the cognitive science community upon a definition of what a cognitive process is, nor of the vehicles of cognition. Or, we could look at the situation a different way, cognitive scientists are pluralistic about the kinds of things they count as cognitive processes and vehicles. For example, classical computationalists take the vehicles of cognition to be symbols that have formal, or syntactic, properties in virtue of which they are processed (Fodor and Pylyshyn 1988). Connectionist vehicles of cognition are not symbolic; instead they are patterns of activation distributed across nodes in a network. Connectionists understand cognitive processes to be algorithms for the spread of activation across the network (Smolensky 1988, 1995).

It is quite natural to be pluralistic about cognitive processes and vehicles; as such, there is no single genuine "cognitive kind".[3] In general, we might specify that a cognitive process is one that involves the manipulation of a cognitive vehicle in the completion of a cognitive task. The classical–connectionist debate demonstrates that there is a plurality of types of manipulations and vehicles. Furthermore, it may be the case that not all cognitive vehicles are representational vehicles. Take, for example, the role of the ambient array in ecological theories of perception (Gibson 1979).

We do have a sense of the cognitive task, to which the notion of cognitive process is surely central. Rowlands gives us a general sense of the cognitive task (2003, p. 161):

> it does seem fairly clear that the notion of a cognitive process is defined, in part, in terms of the notion of a cognitive task. A cognitive process is one that plays a fairly central role in allowing a subject to accomplish a cognitive task.

Quite generally, cognitive tasks are ones such as perceiving the world, remembering things about the world and employing things remembered in making inferences, problem solving and the like (Rowlands 2003). However, a general definition of a cognitive task can easily end up being unhelpfully vacuous. If we define the cognitive task as any task for the completion of which cognition is required, then almost every task will be a cognitive one. I think it is more helpful if we think of cognitive tasks as involving the exercise of particular cognitive capacities such as remembering a date, solving a problem, learning to do something and so on. These are tasks where the exercising of cognitive capacities is directly tied to their successful completion.

Perhaps there is more clarity about the notion of a cognitive system? There are two senses of cognitive system, which we need to distinguish. That of a particular cognitive system – for example the memory system as it might be – and the overall system, of which these specialised sub-systems are parts. However, there is not even agreement on what a cognitive system is in either of these senses. One very general way of thinking about cognitive systems is that they are the mechanisms that underlie the processes involved in remembering, perceiving, learning and reasoning and so on. A theory of cognitive architecture, such as classicism or connectionism, is supposed to specify the nature of these mechanisms. Classicists and connectionists have generally agreed that the mechanisms that underlie cognitive processes are all in the cranium; they endorse the assumption of cognitive internalism. I turn now to look in a bit more detail at the classical and non-classical visions of cognition.

1.3 The classical and non-classical visions

As mentioned in the previous section, cognitive processes and vehicles in cognitive science are understood in both a classical and a connectionist way. Classically cognition has been understood as the processing of representational vehicles. Representations are intentional entities, in that they are directed at something else, this is what it is for them to mean something, or have content. They are complex, in that they can have constituent parts, linguistic representations being a case in point. Call these classical vehicles.

However, not all cognition involves manipulations of vehicles as classically conceived. Neural networks, animate (Ballard 1991), ecological (Gibson 1979) and sensorimotor (O'Regan and Noë 2001) accounts of perception do not involve manipulations of classical vehicles. In neural networks, there are patterns of activation distributed across aggregates of neurons and in ecological theories of perception there are perceptual mechanisms which are directed at environmental variables that afford action. We can identify the directedness of the patterns of activation as well as mechanisms and their affordances, such that they at least have an intentional aspect. What they do not have are the articulate and complexly structured contents of classical vehicles. Call these non-classical vehicles.

So, there is a distinction between classical and non-classical vehicles of cognition and there will be differences in the way they are manipulated. I shall take manipulations of vehicles to be general enough to cover both classical and non-classical cases. I shall take the following definition of a cognitive process as standard throughout the book:

> *A process is cognitive when it aims at completing a cognitive task; and it is constituted by manipulating a vehicle*.

This is a very general definition which does not tie us to a process or vehicle having to be internal or external, or whether vehicles must be representations or have a particular kind of content. I take it to be a working definition rather than a set of necessary and sufficient conditions and I expect each case to be judged on its cognitive merits. I do not think, for example, that switching on a television with a remote control constitutes a cognitive process – care needs to be taken in formulating the cognitive task here.[4] More controversially, I do not think that tapping numbers into a calculator constitutes a cognitive process. Even though there is a clear sense – in which such a manipulation aims at completing a cognitive task – care needs to be taken in formulating the type of manipulation here.[5]

I do not think that internalists will very likely be persuaded by definitions anyway. However, I do think that a careful explanation of the manipulation thesis and types of manipulation in Chapters 5, 6 and 7, will prove more persuasive.

The integrationist takes cognition to be hybrid in that it is made up of classical and non-classical vehicles and processes and that some of these processes and vehicles will be bodily internal and others bodily external. I turn now to outlining the main claims of cognitive internalists (who often call themselves individualists), before moving on to the arguments.

1.4 Cognitive internalism

Cognitive internalists take the distinction between what is inside the head and what is outside of it to be significant. They argue that cognitive processes are located in the head and that consequently the study of cognition should make reference only to what goes on inside the head. Cognitive internalists often refer to themselves as psychological individualists (Stich 1983, Fodor 1987, Egan 1991, Segal 1991, Burge 1986). For example, Stich provides a constraint upon explanation and taxonomy in the cognitive and psychological sciences, which he calls the principle of autonomy:

> The basic idea of the principle is that the states and processes that ought to be of concern to the psychologist are those that supervene on the current, internal, physical state of the organism . . . Any differences between organisms which do not manifest themselves as differences in their *current, internal, physical* states ought to be ignored by a psychological theory. (1983, p. 164) [My italics]

Cognition is largely autonomous of what goes on outside the head, but integrationists argue that cognitive vehicles and processes are, at least, partly constituted by what goes on outside the head. The autonomy principle rules out the possibility of integrationist explanations in cognitive science.

Individualists often use a further methodological constraint, methodological solipsism (Fodor 1980). The psychological and cognitive sciences ought to taxonomise/individuate cognitive vehicles and processes only in terms of their formal properties – this is known as the formality condition. This is because cognitive processes are computational processes and only have access to the formal properties of cognitive states. We get solipsism because the cognitive states in question are taken to be *narrow*, they are states that do not presuppose anything about the external world of the individual who has them (Fodor 1980).

The Individualist has it that cognitive processes, and the cognitive vehicles which those processes apply to, are taxonomised as a kind in terms of their computational properties and causal capacities. If it can be independently shown that computational properties and causal capacities supervene only upon the internal, intrinsic, physical properties of an individual, such as his or her brain states, then internalism about cognitive vehicles and processes follows.

However, the formality condition does not tell us anything about the location of cognitive vehicles, they might well be external. Cognitive integrationists must argue against the claim that the formality condition requires cognitive vehicles to be individuated narrowly. Of course, if there are external cognitive vehicles, then it is clear that they will be given a wide individuation. As such, cognitive integrationists *need* have no quarrel with the formality condition, some cognitive processes may well turn out to be computational;[6] their argument is with the narrow individuation of cognitive vehicles and processes.

Michael Devitt (1990, p. 377) summarises the main claims of individualism:

1 Psychology explains why given certain stimuli at her sense organs, a person behaved in a certain way.
2 Only something that is entirely supervenient on what is inside her skin – on her intrinsic internal physical states, particularly her brain – could play the required explanatory role between peripheral input and output.
3 Environmental causes of her stimuli and effects of her behaviour are beside the psychological point.
4 Cognitive processes and cognitive vehicles must be individuated according to their role within the individual, without regard to their relations to an environment.

I shall now examine the argument from causal capacities and the computational argument, as arguments for the conclusion that cognitive kinds supervene on intrinsic properties of individuals.

1.5 The argument from causal capacities

Fodor puts forward his argument from causal capacities[7] in the second chapter of *Psychosemantics* (1987). A causal capacity is just the capacity to bring about, or cause, an effect. Whilst the argument is primarily designed to show that taxonomy (individuation of kinds) in psychology cannot be by wide content, it is in its own right an argument for cognitive internalism. We shall evaluate it as an argument for cognitive internalism, ignoring the ramifications for theories of content.

The argument really has two crucial steps, the first being that the sciences individuate causal capacities by specifying the relevant intrinsic causal properties a thing has. This claim is false because many sciences individuate by wide causal properties and Fodor himself acknowledges this. The scope of the claim is therefore wrong. Science sometimes individuates causal capacities by intrinsic properties and sometimes by wide[8] properties.

The second step is to insist that causal capacities are supervenient only upon intrinsic properties, because even if wide properties are sometimes used to individuate causal capacities they never themselves constitute causal capacities. This claim is again false across many of the sciences. It drives an intuitive reaction to cognitive integration: how could something bodily external to the organism be part of its causal capacity to do something? In response, I develop the notion of a wide capacity, which is akin to Wilson's wide realisation (Wilson 2004), where a capacity is realised by the organism as part of some wider system that extends beyond the bodily boundaries of the organism. I shall proceed by taking each of these two stages of the argument in turn.

1.5.1 Individuation by intrinsic causal properties

Fodor's first premise is that science taxonomises by causal capacities (what an entity can cause).

> We want science to give causal explanations of such things (events, whatever) in nature as can be causally explained. Giving such explanations essentially involves projecting and confirming causal generalizations. And causal generalizations subsume the things they apply to in virtue of the causal properties of the things that they apply to. Of course. (1987, p. 34)

To follow the argument we will have to go along with this premise and I shall not subject it to scrutiny here.

The second premise can be stated in either a weak or a strong form: the causal capacities an entity has supervene upon the causal properties it has; or the causal capacities an entity has supervene upon the *intrinsic* causal properties it has.

The issue here is simply whether wide properties can be counted as genuinely causal. If we use the weaker version then the possibility that they are is left open. The second option restricts the possibility to intrinsic properties alone. Fodor is inclined to lean towards the weaker version at times:

> In short, what you need in order to do science is a taxonomic apparatus that distinguishes between things insofar as they have different causal properties, and that groups things together insofar as they have the same causal properties. (1987, p. 34)

Let us assume that Fodor is right about this; although some have disagreed,[9] how would we generate cognitive internalism from it? We can generate internalism by narrowing the explanation of a causal capacity to include only intrinsic properties, ignoring wide properties even where they may be relevant. Take the following simple example from the Churchlands:

> A neuron cannot know the distant causal ancestry of its inputs and outputs. An activated neuron causes a creature to withdraw into its shell not because such activation represents the presence of a predator – though it may indeed represent this – but because that neuron is connected to the withdrawal muscles, and because its activation is of the kind that causes them to contract. (1983, p. 305)

Neuron A has the causal capacity to cause muscle contraction, because it has the causal property of spiking at a particular threshold, other causal properties which neuron A has will be irrelevant to this particular causal capacity. So, in the above case the historical property of neuron A – that it has been activated in the presence of a predator – is irrelevant to the causal capacity of muscle contraction. If we are considering how the organism has the capacity to retract muscles, then the causal properties in question will be intrinsic. However, if we want to explain how the creature has the capacity to withdraw into its shell in the presence of predators, then we will need an account of the activation of the neuron in the presence of predators and, therefore, make reference to its wide properties.

If the history of the organism and the environment it is located in are relevant to the explanation of the causal capacity of withdrawing into its shell in the presence of predators, then taxonomy is also by wide properties.

This raises an important issue: the explanation of a causal capacity will be narrow or wide depending upon whether the relevant causal properties are restricted to narrow ones or include wide ones. Such restrictions will depend upon the explanatory context. If I only want to explain the mechanical capacity of neuron A to cause muscles to retract, then of course I will be restricted to intrinsic properties. However, if I am an evolutionary biologist who wants to know how an organism has evolved to retract into its shell in the presence of predators, then I will need to include wide as well as (narrow) mechanical causal properties in my explanation. This does not require that neuron A "knows" anything about its causal ancestry, just that it is part of a wider causal system. The "width" of explanation here depends upon the explanatory project of the science at issue. Physiologists may be interested only in mechanical properties of the organism; evolutionary biologists may be interested in the relationship of the organism to its environmental niche, the phylogenetic history of the species of which the organism is a member and the ontogenetic history of the organism itself. This ought to be a salutary methodological lesson to individualists who espouse the principle of autonomy, or methodological solipsism.

Importantly, cognitive integrationists think that there are wide causal capacities, because some cognitive vehicles and processes are external. In the individualist's terms, some cognitive capacities will be wide, because they rely upon wide causal properties.[10] So we need to distinguish between narrow causal capacities that supervene only on intrinsic causal properties and wide causal capacities that supervene on both intrinsic and wide causal properties. This distinction is important for our consideration of Fodor's way of taxonomising by causal capacities.

Fodor argues that we need a way of determining what kind of causal properties are relevant to determining the causal capacities that a thing has.

> To classify by causal [capacities] is to count no property as taxonomically relevant unless it affects causal [capacities]. But x's having property *P* affects x's causal [capacities] just in case x wouldn't have caused the same events had it not been *P*. (1987, p. 38)

It is important to note that this definition of *affecting causal capacities* is the same for both intrinsic and wide properties. Hence, Fodor ought to adopt only the weak form of the premise – the causal capacities an entity has supervene upon the causal properties it has – but the weak premise is no support for the conclusion that cognitive capacites supervene on intrinsic properties of individuals.

Given this definition of the dependence of causal capacities on causal properties, there is a clear sense in which neuron A would not have caused the muscle contraction if it had not been selected to do so in the presence of predators. The wide causal capacity is dependent upon both intrinsic and wide properties.

It follows that wide and intrinsic properties will be irrelevant to taxonomy by causal capacities, when they do not affect the causal capacities that a thing has. It turns out that Fodor is quite willing to accept that taxonomy is often in terms of wide properties.

> it's patent that taxonomic categories in science are *often* relational. Just as you'd expect, relational properties can count taxonomically whenever they affect causal [capacities]. Thus "being a planet" is a relational property par excellence, but it's one that individualism permits to operate in astronomical taxonomy. For whether you are a planet affects your trajectory, and your trajectory determines what you can bump into; so whether you're a planet affects your causal [capacities], which is all that individualism asks for. (1987, p. 43)

This all seems ecumenical, both intrinsic and wide properties affect causal capacities; therefore, taxonomy will be by both wide and intrinsic properties. As we have seen, Fodor does concede that wide properties can *affect* the causal capacities a thing can have. This concession shows that there is no general motivation for the specific claim that cognitive causal capacities are individuated by intrinsic properties. It shows that the first stage of the argument is really no support for the second. Consequently we shall have to look elsewhere to support the claim that cognitive capacities supervene on intrinsic properties.

1.5.2 Causal capacities supervene on intrinsic causal properties

Given the arguments and concessions of the previous section, it looks clear that Fodor ought to accept the distinction between narrow causal capacities and wide causal capacities. This is because either the causal capacities a thing has are due to its intrinsic properties, or they are due to its wide properties, or they are due to a combination of the two. The final option is all that is required for a wide capacity.

However, Fodor also holds that the causal capacities that a thing has are due only to its intrinsic properties, call this a narrow causal capacity. The general notion that causal capacities can only be narrow is motivated by the general assumption that causal capacities are not constituted by wide properties (Stalnaker 1989).

The crucial caveat here is that wide properties must still affect intrinsic properties if they are to affect causal capacities; this is because, as Stalnaker (1989) puts it, environmental facts do not *constitute* causal capacities. The same metaphysical assumption is behind Fodor's claim that causal capacities must supervene upon intrinsic causal properties; hence he would reject the notion of wide causal capacities.

But he ought to accept the notion, so let us motivate it further by returning to the example of neuron A. Neuron A has the narrow capacity to cause muscle contraction, because it has the causal property of spiking at a particular threshold. However, neuron A also has the wide capacity of causing muscle contractions in the presence of predators. This is because, in the first instance, it has the intrinsic causal property of spiking at a particular threshold. Secondly, neuron A became activated in the ancestors of the organism which had it in the presence of predators and would not have become activated if the predator had not been present.[11] The wide capacity depends on both the intrinsic property of neuron A and the historical and wide properties that cause A to fire in the presence of predators.

Fodor ought to adopt the wide sense of causal capacities, because the organism would not be able to cause itself to retract into its shell in the presence of predators if it did not have both intrinsic and wide causal properties. The individuation of the wide causal capacity depends upon both intrinsic and wide properties. This means that some wide properties are properly causal properties in determining the causal capacities of a thing. According to Wilson, there certainly are biological cases where wide properties function as explanantia (as causes):

> For example, being a Mother, being unemployed, being a member of a particular species, being a planet, being located in a magnetic

> field, and occupying a relatively specific ecological niche are all relational properties that different properties can have in particular instantiating circumstances, each of which, when coinstantiated with the appropriate properties, enables an entity to bring about particular effects. With respect to causal efficacy, some relational properties are *just like* paradigmatic intrinsic properties. . . . (Wilson 1995, p. 124)

Explanations that make reference only to narrow causal capacities allow us to refer only to the intrinsic causal properties of neuron A. This would be the construal that accords with the principle of autonomy.

The wide construal is dependent upon neuron A coming to have the biological function that it does. This would be a matter of selective pressures over generations, or a function acquired by learning or re-inforcement. In this case the explanation is by the wide causal capacity that the organism has which is dependent upon its historical relations to its environment (its wide properties). Therefore, the biological function of the organism to contract muscles in the presence of predators is dependent upon relations to the environment and it would not have the causal capacity of contracting muscles in the presence of predators if it did not have this wide property. All Fodor asks for a causal capacity to be dependent upon a causal property is that without that property the causal capacity would not be able to bring about its effect. So by Fodor's own definition, wide causal capacities are on the cards.

It follows that internalists, such as Fodor, ought to concede that there are both narrow and wide causal capacities. They should make this concession because the kind of explanation that a science requires will determine whether wide or narrow causal capacities are used as part of that explanation. However, if Fodor, or any other individualist, was to endorse the use of wide causal capacities, he would have to accept that they are constituted by both wide causal properties and intrinsic causal properties. Therefore, the cherished metaphysical assumption that only intrinsic properties can be truly causal cannot be used to motivate psychological individualism.

This accords with Wilson's notion of wide realisation. A wide realisation is one where a system extends beyond the boundary of the individual which has it (Wilson 2004, p. 112). The point can be made in terms of dispositions; Wilson points out that many dispositions in the physical sciences are relational (wide), whether or not a liquid is an acid is one such example. Being a particular chemical substance involves a disposition to have certain effects on other substances and the having of these dispositions is determined by the nature of those other substances (Wilson 2004).

> The very presence of the disposition, not just its manifestation, involves the physical configuration of the world beyond the bearer of the disposition. . . . This point is reflected in standard definitions of an acid – for example, as a proton donor, or as an electron-pair acceptor. For whether a substance with a given physical structure has the disposition to donate protons or to accept electron pairs depend upon facts about the broader chemical system in which that substance exists. If just these facts were different, a liquid that is actually an acid could lose this disposition, and could do so even were its chemical composition to remain unchanged. (Wilson 2004, pp. 125–6)

Fodor has to concede that there are wide causal capacities to be found even in the physical sciences, because some wide properties act as causal properties. However, he does not think that there can be wide psychological causal capacities. This is because, in the psychological case wide properties affect causal capacities only by affecting the intrinsic properties of an individual, he and Stich are as one on this. So I turn now to dealing with the third premise.

Premise 3: *In the psychological case the relevant causal properties supervene upon local neural structure*.

- *Methodological point*: Categorization in science is characteristically taxonomy by causal capacities. Identity of causal capacities is identity of causal consequences across nomologically possible contexts.
- *Metaphysical point*: Causal capacities supervene on local microstructure. In the psychological case, they supervene on local neural structure. (Fodor 1987, p. 44)

So far we have been dealing with the methodological point, but the metaphysical point is crucial. The causal capacities a thing has (what effects it can cause) supervene on local microstructure (in the case of psychology on neural structures). Fodor does not allow that causal capacities supervene upon wide properties and this coheres with the principle of autonomy. But why does Fodor think this?

> We abandon this principle at our peril; mind/brain supervenience (/identity) is our only plausible account of how mental states could have the causal powers that they do have. (Fodor 1987, p. 44)

The intuition at work here is obvious; only intrinsic properties are really causally efficacious, but this intuition is mistaken, as we have already seen. Consequently, the narrow version of the second premise of Fodor's argument is false, because causal capacities in general do not supervene exclusively on intrinsic properties. So, it looks increasingly likely that premise 3 will be false as well.

Returning to our example, the narrow causal capacity of neuron A to contract muscles in the presence of predators is dependent upon local neural structure. Neuron A is connected to muscles and must transmit electrochemical messages along these connections, before any contraction is possible. The wide causal capacity of the organism to retract into its shell in the presence of predators depends upon relational and historical properties of the organism, as well as intrinsic ones. The wide causal capacity of the organism supervenes upon both its intrinsic properties and its wide properties. The question is, are there any cognitive capacities that are wide in this sense?

Fodor thinks this is unlikely because there must be a mechanism to bring about any causal effects and the mechanism must be physical. In the psychological case, although relational properties will affect causal capacities and individuation will often be wide, the psychological mechanism specified must supervene on a physical/neural mechanism. Fodor stipulates that you cannot affect causal capacities of mental states, except via the physiology of the organism:

> you *can't* affect the causal powers of a person's mental states without affecting his physiology. . . . God made the world such that the mechanisms by which environmental variables affect organic behaviours run via their effects on the organism's nervous system. (Fodor 1987, p. 40)

Wide cognitive capacities might be like biological capacities that have been selected for, as in our simple example of neuron A; but they might be like the causal capacities of a planet. A planet has its causal capacities because of the present relations it has to other objects and forces in its environment. The planet's velocity is dependent on both its intrinsic properties and its wide properties. There is a mediating law or mechanism to allow the continual causal influence of the environmental factors. It follows that the relations a cognitive system stands in at any particular time may affect the causal capacities that it has.

For example, sensorimotor contingency and ecological accounts of perception depend upon wide properties. They depend upon continuous causal interaction between the ambient array of light, the movement of the body and the brain. What we perceive depends upon the continuous causal influence of environmental factors.

In Chapters 4, 5, 6 and 7, I provide a comprehensive range of examples of wide cognitive capacities. Therefore, Fodor's metaphysical claim that cognitive causal capacities supervene only on intrinsic properties is false.

Fodor holds that as long as a causal property affects causal capacities, then it is taxonomically relevant and this allows for individuation by wide properties. However, he assumes that the relevant psychological causal capacities supervene on the brain and that if wide properties are to affect causal capacities they must do so by affecting intrinsic properties of brains. This is because, only intrinsic properties of neural mechanisms are truly causally efficacious. The assumption of local supervenience of causal capacities on intrinsic properties is not global, because wide properties are often causally efficacious. Therefore, there are wide causal capacities.

It is then an open question whether or not there are wide cognitive capacities, but there is nothing in the argument from causal capacities which precludes them, except for an adherence to mind–brain supervenience. The evidence is stacking up against mind–brain supervenience; cognitive integrationists provide many examples of wide cognitive capacities, for which I direct the reader to Chapters 4, 5, 6 and 7.

1.6 First stage of the computational argument: Methodological solipsism

As we saw in the first section of this chapter, a methodological constraint follows quite naturally from mind–brain supervenience:

that the states and processes that ought to be of concern to psychological explanation are those that supervene on the current, internal, physical state of the individual. Methodological solipsism gives us a way of making this methodological constraint hold in computational cognitive science. The computational version of methodological solipsism depends upon what Fodor (1980) calls the "formality condition": computational processes have access only to the formal properties of representations and not the semantic ones, where semantic properties are taken to be wide properties. Solipsism is supposed to be derived from the formality condition, because formal properties, inevitably, supervene upon intrinsic properties of an individual, hence mind–brain supervenience holds.

Yet we can easily see how the formality condition is quite consistent with cognitive integration on the assumption that the manipulations of external vehicles have access to the formal properties of those vehicles.[12] Given that the vehicles are *externally located* and that they are manipulated externally, solipsism cannot follow. Therefore, cognitive science does not individuate cognitive states and processes narrowly (without reference to the environment). Consequently, even in computational cognitive science individualism does not hold. I shall now look at how methodological solipsism is argued for.

Putnam claims that the internalist conception of mind is based on an assumption of "methodological solipsism" which is derived from Descartes.

> This assumption is the assumption that no psychological state, properly so called, presupposes the existence of any individual other than the subject to whom that state is ascribed. (Putnam 1975, p. 10)

Descartes initiated a theoretical divide between mind and world, where an internal subject is related to an external world via a perceptual interface. This, of course, ultimately leads to a methodology for studying mind based upon the divide. Descartes was also an early exponent of the representational theory of mind (henceforth RTM): the internal mind is related to the external world by representations, ideas, causally connected via perceptual interfaces to the external environment.

Contemporary exponents of RTM, in cognitive science, claim that the mind supervenes on the brain. The representations and processes over those representations can safely be characterised as physical or supervening on the brain. This is a way of thinking about the mind that makes it "self-contained," from its environment. This is to say that the environment makes no essential difference *to the state of mind we are in* (Fodor 1980). It relies upon the thesis of the local supervenience of computational states on brain states, and the claim that the processing of representations has no immediate access to the external environment.

The methodological solipsist takes this Cartesian claim seriously. He distinguishes between a *rational* psychology and a *natural* psychology (Fodor 1980). The rational psychologist takes seriously the Cartesian injunction that the way the world is makes no essential difference *to the state of mind we are in*. (Fodor 1980) Naturalists, by contrast, focus on the fact that the organism is embedded in an environment and what the relevant organism–environment interactions are (Fodor 1980, p. 487). According to the methodological solipsist, we should only do rational psychology, because natural psychology is too difficult. This is so because natural psychology needs to explain organism–environment relations.

The natural psychologist has to provide law-like generalisations covering each of these relations by specifying the relation between an organism and an object in its environment, such that the organism is thinking about that object. As such the theory would need to

> define its generalizations over mental states on the one hand and environmental entities on the other, it will need, in particular, some canonical way of referring to the latter. (Fodor 1980, p. 496)

Natural psychology would depend upon other sciences for the characterisation of the relation and the canonical description of the objects the organism is related to. This marks out the difficulty in doing natural psychology. Take the following example used by Fodor (1980, p. 496):

1. Salt is the object of what "Granny desires to put on her herring."
2. We rely on chemistry to give us a canonical description of salt as NaCl.

3. This requires that for all the other objects of thought there is a canonical description of the object forthcoming from one of the sciences.
4. This isn't true.
5. Therefore, natural psychology is too difficult.

Fodor's complaint is that natural psychology makes psychological methodology dependent upon other sciences. He takes this to be pernicious, because

> (a) that we don't know relevant nomologically necessary properties of most of the things we can refer to (think about) and (b) that it isn't the linguist's (psychologist's) job to find them out. (Fodor 1980, p. 496)

Therefore, we should restrict ourselves to a rational psychology. This conclusion applies not just to causal theories of content, but to any psychological theory that makes use of organism–environment relations in general, such as Gibson's ecological theory of perception. Of course, Fodor has changed his mind on this, he now endorses a natural psychology of content where the meaning of mental representations is determined by causal relations to environmental objects. Hence, natural psychology is on the cards even for Fodor.

We could restrict rational psychology to computational psychology where methodological solipsism depends upon the formality condition. Computational processes (and mental processes if they are computational) are both *symbolic* and *formal*:

> They are symbolic because they are defined over representations, and they are formal because they apply to representations, in virtue of (roughly) the *syntax* of the representations. (Fodor 1980, p. 486)

Syntactic operations are defined over a symbol's intrinsic physical properties, such as its shape. These operations are not sensitive to any of the semantic properties a symbol may have, that depend upon their relations to the environment. Fodor goes on to say that the

> Formality condition connects with the Cartesian claim that the character of mental processes is somehow independent of their

> environmental causes and effects. If mental processes are formal, then they have access only to the formal properties of such representations of the environment as the senses provide. (1980, p. 488)

However, the formality condition does not establish methodological solipsism, because the computational vehicles and processes might be environmentally located. Only if we assume that computational processes and vehicles supervene upon internal neural structures can we make the formality condition count in favour of methodological solipsism. This follows if we take the doctrine of the local supervenience of the mind upon the brain to be true, but we have so far seen no good reason for accepting it (see the previous discussion on wide causal capacities). Let us try an example, from Kim Sterelny (1990, p. 35), to help clarify the claim.

My seeing a tree causes a visual representation, A, of that tree in my visual cortex. The visual representation A causes the representation, B, "that is an oak tree." Whilst the distal environment – the oak tree – caused A, B was caused *only* by A (even if A had been produced by a hologram). The causal ancestry of A is irrelevant to the tokening of B; this could only be the case if B comes about because of features of A which are narrow, or internal to the system. Computational processes do not have access to these relational properties of A.

As with neuron A (from the previous section), representations A and B are merely sequences in a long, albeit hugely simplified, causal chain. A different way of viewing the sequence would be to view the environment and "internal" representations as an integrated system, where causal interactions go both ways. The question though is why we should not do this?

As we have seen, the formality condition is consistent with the representations in question being located in the environment, because it is a condition only on the properties that computational processes have access to. The formality condition says nothing about mental representations and mental processes having to supervene upon the brain, this is assumed by the internalist.

As we shall see, in the next section, this is the main problem with the computational argument for internalism. The internalist asserts that the external physical and social environment of an individual have no constitutive relevance to that individual's mind. In methodological mode, psychological states are taxonomised in such a way that the nature of the external world of the individual is irrelevant. However, if vehicles of cognition supervene on the environment, then we must include the environment as well as the discrete individual in our cognitive explanations. Narrow individuation would fail in these cases, because psychological states would have to be individuated in such a way that the world is relevant to the state of mind we are in. Therefore, methodological solipsism restricts the explanatory framework of computational cognitive science in a pernicious way.

The computational argument, just like the argument from causal capacities, relies upon the truth of mind–brain supervenience and its truth is assumed.

1.7 The second stage of the computational argument: Integrated computational systems

Robert Wilson (1995, p. 64) provides a simple version of the computational argument for individualism:

(A) The sciences of cognition taxonomically individuate mental processes only qua computational processes.
(B) The computational states and processes that an individual instantiates supervene on the intrinsic states of that individual. Therefore,
(C) The sciences of cognition individuate states and processes that supervene on the intrinsic physical states of the individual who instantiates those states and processes.

This argument is unsound, if premise B is false. Cognitive integrationists think that it is false and that the conclusion does not follow. Premise A is quite possibly false as well if the notion of computation is too restrictive, see above for a discussion of pluralism about processing and vehicles in cognitive science. The computational arguments I will look at are based on a classical – that is not a connectionist – conception of computational processes and vehicles.

What we now need to do is find out why, if the cognitive system is a computational system, it is individualistic? Let us begin by asking what the computational theory of mind is committed to, before looking at Segal (1991) and Egan's (1992) versions of the argument.

A classical computationalist account of mental processes entails that there is a causal sequence of thoughts – A–B–C–D. Each thought is an explicit symbol token and each antecedent token is causally responsible for each consequent token.

A. Fred believes it is too hot in here.
B. Fred notices that the fire is on full.
C. Fred hypothesises that turning down the fire would reduce the heat in the room.
and
D. Fred desires to turn down the fire.
Causes:
E. Fred turns the fire down

Each thought is causally responsible for the next and this is only possible if

1. Each thought is explicitly tokened.
2. The causal relations between thoughts mirrors the semantic relations between thoughts.

If B was "Fred noticed that the sun was shining in through the window," then thoughts C and D would be different thoughts, in virtue of the content of B. Thinking involves a causal sequence of explicit thoughts, where the sequence is determined by what the thoughts mean.

We have seen what a sequence of thoughts is, but how do these thoughts get their causal capacities? They do so by being computational states. We can generally understand this in the following way. A computing machine is a device which has the following features:

- It contains media in which symbolic representations can be stored. These symbols can be arranged into expressions in virtue of their syntactic structure. The symbols mean something, they are interpreted.
- A computer can differentiate between symbols, via distinctions in their syntactic "shape."
- The computer can cause the tokening of new symbols.

- The causal processes that govern what new symbols will be tokened are dependent upon the syntactic form of the symbols already stored in the machine.

Imagine a computer which has the numbers 2 and 3 stored at locations A and B. A new tokening of a representation of a sum of these two numbers will be caused to occur at location C. The symbol tokened at C depends upon the symbols at A and B, but the new tokening at C is strictly dependent upon the syntactic form of the symbols at A and B and not upon their interpretations.

The interesting thing is that the semantic properties of the symbols play no causal role in the process. However, the claim is that the semantic distinctions between the symbols are preserved by the syntactic distinctions between the symbols, and the syntactic type of the symbol determines its causal role in a process.

Semantic properties can have a causal role indirectly because any semantic differences between symbols are reflected in their syntactic distinctions: for any two symbols S and S*, if they differ with respect to their semantic properties then, according to the computational theory, they must also differ with respect to their syntactic properties.

We can see how in a sequence of thoughts, consequent thoughts are dependent upon antecedent thoughts, because, as with the above example, the syntactic properties of the antecedent representations determine which consequent representations get tokened; and the semantic relations between antecedent and consequent representations are mirrored by the syntactical relations between the antecedent and the consequent representations.

Wilson sites Egan as an individualist who thinks that it follows from this characterisation of the computational theory of mind that it is individualistic in the way that it conceives of mental states and mental processes.

> "Symbols are just functionally characterised objects whose individuation conditions are specified by a *realization function* f_R which maps equivalence classes of physical features of a system to what we might call 'symbolic' features. Formal operations are just those physical operations that are differentially sensitive to the aspects of symbolic expressions that under the realization function f_R are specified as symbolic features. The mapping f_R allows a

> causal sequence of physical state transitions to be interpreted as a *computation*.
>
> Given this method of individuating computational states, two systems performing the same operations over the same symbol structures are computationally indistinguishable." From this she concludes that "if two systems are molecular duplicates then they are computational duplicates. Computational descriptions are individualistic – they type-individuate states without reference to the subject's environment or social context." (Egan 1992, p. 446)

Wilson notes (1995, p. 68) that Egan's conclusion only follows if we equate the computational system with an individual subject. This begs the question against cognitive integration, or Wilson's "wide computationalism".[13] The claim that mental states and mental processes must supervene on intrinsic properties of subjects does not follow from the formality condition (as we saw in the previous section). This is because the formality condition states only that, *Computational processes have access only to the formal properties of representations, such as their size and shape*. It does not state that representations and processes *must* supervene upon intrinsic properties of subjects, they could, in principle, be spread out over subject and embedding environment. So, that line of argument is now closed to the computational individualist.

Segal (1991) argues for an individualist reading of computational systems, by invoking the notion of an integrated computational system:

1. The representational states that a computational system is in are determined by the intrinsic properties of the system.
2. Computational systems are integrated.
3. The computational system is integrated in virtue of the supervenience base of properties, which the representational states and computational states of the system supervene upon.
4. The intrinsic properties of the supervenience base are the determinants of the representational states that the system is currently in as well as the state changes which the system is capable of.

As we have just noted, the first claim cannot be made via the formality condition. The individualist may just wish to *stipulate* that formal properties are intrinsic properties, because they supervene on intrinsic properties of individual brains. As such, this would preclude the extension of the computational system beyond the discrete individual. However, this is an entirely empirical claim, and if there are vehicles in the environment that are processed in terms of their formal properties, then these properties are not intrinsic to the individual.

Let us turn to the second point about integration. Internalists think that a computational system is physically integrated when all its components have a causal influence over one another. Cognitive integrationists accept that components must have a causal influence over one another, but reject that they *must* be located exclusively in the body.

The internalist is worried that if a cognitive system had some of its mental representations spread out beyond the boundary of the discrete individual, how would the cognitive processes which supervene on neural mechanisms get access to them? Computational processes would have remote access to these symbol structures, but this commits us to action at a distance, which is absurd. The simple response is that not all cognitive processes supervene on the brain, some processes are constituted by bodily manipulations of external representations and no action at a distance is implied by that.

The computational argument for internalism fails again. It leaves open the possibility that the cognitive integrationist could provide examples of external and internal cognitive processes having a causal influence over one another. This would be to show that the "cognitive system" is integrated even if some vehicles of cognition and the processing of those vehicles occurs outside the head. If the integrationist is right, as I hope to show, then the supervenience base of properties will include wide properties as well as intrinsic properties. Segal is willing to wager an empirical bet with the integrationist:

> Individualism is the thesis that the representational states of a system are determined by intrinsic properties of that system. It seems likely that whole subjects (or whole brains) make up large, integrated, computational systems. Whole subjects plus embedding environments do not make up integrated, computational systems. That is one reason why individualists draw the line where

> they do: the whole subject is the largest acceptable candidate for the supervenience base because it is the largest integrated system available. (Segal 1991, p. 492)

This, like the other reasons given in this chapter, turns out not to be a good enough reason to be an individualist/internalist. In the second part of this book we will see numerous examples of the integration of internal and external vehicles and processes in the completion of cognitive tasks.

What the arguments for individualism show is that internalism about mental states and mental processes is motivated by the assumption of the local supervenience of the mind upon the brain. Cognitive scientists do not always respect local supervenience in their taxanomic and methodological practices. Like the metaphysical assumption that only intrinsic properties are causal, it is a bad assumption. There is nothing that precludes cognitive scientists breaking with the autonomy principle in practice – and they do so on a regular basis.

1.8 Conclusion

We have seen that cognitive internalism, as it is expressed in psychological individualism, relies upon the assumption of the local supervenience of the mind upon the brain: all cognitive properties, states and processes *must* supervene upon intrinsic properties of the brain. We have seen that there are two arguments for this kind of internalism: the argument from causal capacities and the computational argument for individualism in psychological explanation. Fodor hoped to show that causal capacities supervene *only* upon intrinsic properties of individuals, in this sense causal capacities are narrow. However, his argument allows for wide causal capacities and leaves open the possibility that there are wide cognitive capacities.

The computational argument assumes that local supervenience is a necessary condition for the computational theory of mind. Computational processes have access only to syntactical properties of representations, mental representations supervene upon the brain, and therefore computational processes must supervene upon the brain. This makes sense if we are worried about cognitive processes extending out beyond the boundary of the skin. This worry only holds if the internal and the external components of a cognitive system are not integrated, because its components must have a causal influence over one another. The computational argument also relies upon an empirical bet. As Segal reminds us, there just are no psychological theories that treat whole subjects plus embedding environments as integrated cognitive systems. This bet has been taken up and I know where my money lies.

Both arguments are refuted when we find evidence of cognitive causal capacities supervening upon wide properties and psychological theories which treat whole subjects and embedding environments as integrated cognitive systems. More and more of this kind of evidence is emerging from the cognitive sciences. This casts doubt on the reliance on mind–brain supervenience. We ought to have good reasons for accepting the local supervenience of the mind upon the brain, it should not merely be assumed as a dogma. In the next chapter, I will look at some definitions of externalism and outline the dynamical approach to integration.

Index